Annotating Medical Image Data

Grünberg, K ; Jimenez-del-Toro, O ; Jakab, András

DOI: https://doi.org/10.1007/978-3-319-49644-3

Chapter 4
Annotating Medical Image Data

Katharina Grünberg, Oscar Jimenez-del-Toro, Andras Jakab, Georg Langs, Tomàs Salas Fernandez, Marianne Winterstein, Marc-André Weber and Markus Krenn

Abstract This chapter describes the annotation of the medical image data that were used in the VISCERAL project. Annotation of regions in the 3D images is non-trivial, and tools need to be chosen to limit the manual work and have semi-automated annotation available. For this, several tools that were available free of charge or with limited costs were tested and compared. The GeoS tool was finally chosen for the annotation based on the detailed analysis, allowing for efficient and effective annotations. 3D slice was chosen for smaller structures with low contrast to complement the annotations. A detailed quality control was also installed, including an automatic tool that attributes organs to annotate and volumes to specific annotators, and then compares results. This allowed to judge the confidence in specific annotators and also to iteratively refine the annotation instructions to limit the subjectivity of the task as much as possible. For several structures, some subjectivity remains and this was measured via double annotations of the structure. This allows the judgement of the quality of automatic segmentations.

Source code is available at:
https://github.com/Visceral-Project/annotationTicketingFramework

K. Grünberg · M. Winterstein · M.-A. Weber
University of Heidelberg, Heidelberg, Germany
e-mail: katharina.gruenberg@med.uni-heidelberg.de

O. Jimenez-del-Toro (✉)
University of Applied Sciences Western Switzerland (HES–SO), Sierre, Switzerland
e-mail: oscar.jimenez@hevs.ch

A. Jakab · G. Langs · M. Krenn
Medical University of Vienna, Vienna, Austria
e-mail: georg.langs@meduniwien.ac.at

T. Salas Fernandez
Agencia D'Informció, Avaluació i Qualitat en Salut, Catalonia, (GENCAT), Spain

A. Hanbury et al. (eds.), *Cloud-Based Benchmarking of Medical Image Analysis*, DOI 10.1007/978-3-319-49644-3_4

4.1 Introduction

Since during clinical routine, only a very small portion of the increasing amounts of medical imaging data are used for helping diagnosis, the VISCERAL (Visual Concept Extraction Challenge in Radiology) project aimed at providing the necessary data for developing clinical image assessment algorithms. An objective was to conduct Benchmarks for identifying successful computational strategies. The VISCERAL project developed a cloud-based infrastructure for the evaluation of detection, analysis and retrieval algorithms on large medical image datasets [8, 9]. VISCERAL organized benchmarks to exploit and compare multiple state-of-the-art solutions designed for image segmentation, landmark localization and retrieval [13]. The VISCERAL Anatomy Benchmarks focused on automatic identification, localization and segmentation of organs in image volumes. An anatomical reference annotation dataset, the Gold Corpus, was created for these Benchmarks using CT (computed tomography) and MRI (magnetic resonance imaging) volumes annotated with up to 20 organs and 53 landmarks each.

One goal of the VISCERAL project was to create a large dataset containing high-quality expert annotations in medical imaging data (i.e. organ segmentations, landmark localizations and lesion annotations). For this purpose, various manual and semi-automatic segmentation tools were evaluated in the search for fast and effective 3D annotation software interfaces that can reduce the time spent and workload of the radiologist making the manual segmentations and annotations of the structures. A ticketing framework was also developed to facilitate the management of multiple annotation types, the distribution of annotation tickets to multiple annotators and the implementation of a quality control procedure to ensure consistent annotation quality across annotators. This chapter describes the two selected annotation tools, the framework that was built to monitor and distribute annotation tickets, the typical life cycle of an annotation ticket, detailed annotation guidelines for the annotators and the procedure of determining the inter-annotator agreement.

4.2 3D Annotation Software

With the ever-increasing amount of patient data generated in hospitals and the need to support a patient diagnosis with these data, computerized automatic and semi-automatic algorithms are a promising option in the clinical field [6]. An initial step in the development of such systems for diagnosis aid is to have manually annotated datasets that are used to train and implement machine-learning methods to mimic a human annotator. The manual segmentation of the patients' 3D volumes is commonly used for radiology imaging in order to separate various structures in the images and allow processing tissue of the structures separately. Manual segmentation, on the other hand, demands an intensive and time-consuming labour from the radiologists.

Variation and errors in the segmentations are common, depending on the experience of the annotator [12].

Several tools have been developed for the manual and semi-automatic segmentation of anatomical structures and annotation of pathologies present in medical imaging [3, 4, 10, 12, 14]. The implemented segmentation methods range from simple manual outlining in 2D cross sections to more elaborated solutions, like deformable registration that finds spatial correspondences between 3D images and a labelled atlas [12]. An important feature of manual or semi-automatic segmentation methods is that they assist the radiologists in the final decision of the resulting 3D structures [7]. Some of these tools are added to application frameworks that provide visualization and image analysis for an integral medical image computing experience.

An objective of the VISCERAL project was to take advantage of effective user-friendly annotation tools that can reduce the time necessary for annotations and segmentations in a multimodal imaging dataset (MRI, CT). Visualization frameworks are also available that reduce the time to develop new applications through the combinations of algorithms, which is usually faster than writing code [1, 2]. Various available tools were explored for the selection of the tool used for the annotation tasks in the VISCERAL project. The selected annotation tool had to make annotations in CT and MRI images acquired with a variety of scanners and in different MR sequences such as T1 weighted and T2 FLAIR, and with a resolution of the annotated voxels of 1 cm or lower. To ensure sustainability, tools with at least a minimum of support were preferred. These requirements further include adaptability of the included segmentation method to overcome the differences in image contrast and resolution in the dataset.

A brief description of the medical annotation functionality of the evaluated tools is presented in the following sections. The criteria used for selecting the definitive tool are also mentioned. Finally, the description of the methods and use of the proposed tools are discussed. The selected tools allowed the radiologists to segment 20 relevant structures of 15 organs in the human body, identify up to 53 landmarks and detect pathological lesions in full-body patient scans. Both individual voxels and homogeneous regions were labelled in the 3D volumes of the dataset. Medical raw data to be annotated were in the DICOM (Digital Imaging and Communications in Medicine) format. The raw DICOM files were converted into NIfTI (Neuroimaging Informatics Technology Initiative) format, because this is a widely used and accepted format that significantly reduces the file size in case of large 3D data. Both the image data and the resulting volume annotations were used in the NIfTI format.

4.2.1 Evaluation Criteria

A list of ten criteria was defined for the comparison between the tools. The goal was to evaluate their main functionality applied to the annotation tasks needed for the creation of the VISCERAL Gold Corpus. The criteria chosen to compare the available annotation tools were as follows:

- ability to perform 3D annotation in CT and MRI volumes;
- flexibility to segment different structures and points of interest;
- user-friendly segmentation method;
- optimal visualization of the medical images and segmentations;
- effectiveness of the segmentation;
- interactive user corrections in the semi-automatic output;
- time spent in a complete structure segmentation;
- adaptability to data obtained from different scanners and image contrasts;
- format of the output segmentations;
- upgrading of the tool with minimum technical support during and after the project has ended.

4.2.2 Reviewed Annotation Tools

For the selection of the VISCERAL annotation tool, the visualization and application frameworks that are already available free of charge were evaluated. The frameworks had to contain a semi-automatic segmentation tool that could reduce the time required for making manual annotations of 3D structures and points of interest. Six frameworks with no license fees: GeoS [4], ITK-SNAP [14], ImageJ, MeVisLab, MITK and 3D Slicer [11], were included in the study. Some Web-based applications with annotation functionality available such as [10] are limited to a specific application or image analysis type, making them unfit to be used for the VISCERAL project multistructure annotation task. Other available frameworks such as SciRun,[1] Osirix[2] and Volview[3] were also reviewed but were discarded early in the selection process.

4.2.2.1 3D Slicer

3D Slicer[4] is a module-based software where each module performs a particular image processing task. There are two modules that can be useful for segmenting and annotating medical 3D images. The first is called Simple Region Growing Segmentation and it is based on intensity statistics. After choosing a desired number of fiducials in the region of interest, it applies ITK filters for curvature flow and connected confidence producing a 2-class segmentation. The segmentations can be improved by increasing the number of iterations, the multiplier and the neighbourhood radius options. More than one fiducial or seed is allowed for refinement of the output. The other module included is EMSegment Easy that performs a quick intensity-based image segmentation on MRI. The user defines the volumes to be

[1] http://www.sci.utah.edu/cibc-software/scirun.html.
[2] http://www.osirix-viewer.com.
[3] http://www.kitware.com/opensource/volview.html.
[4] http://www.slicer.org.

segmented, specifies the number of structures and can add additional subclasses of the structures. Samples are taken from the structures of interest to define the intensity distribution and the weighting of a node in the tree. Once the algorithm is run, the target images are segmented and the label map with corresponding statistics is returned [11].

4.2.2.2 GeoS

The Microsoft medical image analysis project InnerEye focuses on the automatic analysis of the patients' scans. Its annotation tool GeoS[5] has an algorithm to efficiently segment 3D images using a geodesic symmetric filter with contrast-sensitive spatial smoothness. Its behaviour is comparable to that of graph cut algorithms but with a much faster implementation. The segmentation method is based on a generalized geodesic distance transform (GGDT). A geodesic distance map is initialized from a soft seed mask. The seed region is determined interactively. It uses different brush strokes to quickly indicate a foreground object and the background that surrounds it. In this matter, geodesic distance is described as the distance between two points in an image that takes into account image contents such as intensity gradients. One of the most sought-after requirements is edge-sensitivity whereby the image processing system is able to change its behaviour depending on the local image contrast. This tool is able to perform contrast-sensitive image editing or processing. It shares some of the image processing tasks unifying previously diverse image techniques in such a manner that at least some processing may be shared so that computational resource requirements can be reduced.

4.2.2.3 ITK-SNAP

ITK-SNAP[6] is a software application that provides a set of tools for segmenting medical images' volumetric data. The software provides both an algorithm referred to as "Snake evolution" and a visualization interface for 3D image segmentation. The contour evolution on which its algorithm is based uses the image gradient information and the global intensity to expand or constrain the contour with respect to user given seed points. It provides a segmentation pipeline in three steps with three modifiable parameters that influence the output of the segmentation: balloon force, curvature force and advection force. These three parameters regulate the region growing expansion of the segmentation and the smoothness of the output borders. The framework includes a wizard for image upload and also a polygon tool that allows the user to perform freehand annotation. The freehand annotation tool can be expressed in either a continuous curve or piecewise linear with an adjustable segment length.

[5] http://research.microsoft.com/en-us/projects/geos.

[6] http://www.itksnap.org.

4.2.2.4 ImageJ

ImageJ[7] contains several image segmentation algorithms based on intensity range thresholds. In particular, the Robust Automatic Threshold Selection (RATS) performs a threshold on previously established regions using a recursive quad-tree architecture. It calculates the sum of the original voxels weighted by the gradient pixels. Other plug-ins such as the watershed algorithm are available for segmenting images but they mostly rely on histogram thresholding and Gaussian modelling of the intensity values in the images, which can provide an initial estimate but has to be completed by the user with freehand 2D slice-by-slice manual annotation.

4.2.2.5 MeVisLab

MeVisLab[8] is an integrated development environment with a modular framework that allows developing image processing algorithms and visualization and interaction methods. It is possible to create an end-user application with a network composed of modules based on Open Inventor scene graphs, OpenGL, ITK, VTK and SDK. It supports DICOM files as well as NIfTI formats. Conversion of one format to the other is also included within the available modules. Although there are few segmentation algorithms outside those available in the ITK and VTK libraries, the user can use LiveWire combined with freehand manual annotation on a slice-by-slice basis. LiveWire is a graph cut algorithm where the user can adjust the gradient, Laplacian and directional weighting. There is also a "bulge" module that can easily bend, expand and contract manual annotations with the mouse.

4.2.2.6 MITK

The Medical Imaging Interaction Toolkit (MITK)[9] was created as a software system for development of interactive medical image processing software. It implements both ITK and VTK libraries but also offers additional development and interactive features of its own like 3D-synchronized multiviewer layout. It contains various segmentation methods based on threshold functions such as the Otsu segmentation where it is possible to define a number of regions based on a Gaussian modelling of the intensity value image histograms. It is also possible to apply a region growing algorithm with a user-given seed. The framework only allows one seed per region and freehand wiping, correction and filling of the created segmentation. Another option when manual 2D slice segmentations are available is to interpolate the missing slices and create a surface of a structure of interest.

---

[7] http://rsb.info.nih.gov/ij.

[8] http://www.mevislab.de.

[9] http://www.mitk.org.

Table 4.1 Report on the evaluation criteria for each of the frameworks or annotation tools

GeoS	ImageJ	ITK-SNAP	MeVisLab	MITK	3DSlicer	Evaluation Criteria
+	+	+	+	+	+	Annotation on 3D volumes (CT and MRI)
+	-	+	+	+	+	Semi-automatic segmentation method different shapes flexibility
+	-	o	+	o	+	Segmentation user-friendly usage
o	o	+	+	+	+	Optimal image visualization of segmentation
+	-	o	+	+	+	Effectiveness in VISCERAL data of the segmentation method
+	+	+	+	+	+	Interactive output corrections
+	o	+	-	o	o	Semi-automatic algorithm time reduction vs. manual annotation
+	-	o	-	+	+	Local image contrast flexibility
+	o	o	-	+	o	Output annotation format
+	-	-	-	o	o	Upgrading of the tool

+: Satisfactory o: Insufficient -: Missing

4.2.3 Tool Comparison

In this section, we discuss how each of the considered tools satisfies the evaluation criteria. Table 4.1 summarizes the evaluation of the frameworks and annotation tools for each of the evaluation criteria listed in Sect. 4.2.1.

4.2.3.1 Compatible 3D Annotation on CT and MRI Volumes

ITK-SNAP and ImageJ ask for a greyscale or RGB image when they upload and ITK-SNAP uses a wizard for loading a file. Intensity values with an intensity precision larger than 16-bit are approximated. MeVisLab can upload DICOM volumes and Analyze-formatted files but NIfTI files are not supported. Both MITK and 3D Slicer can upload a wide range of different image formats and contain converting format functions. GeoS does not support DICOM files but works with NIfTI files as well as other image formats such as Analyze and Tagged Image File Format.

4.2.3.2 Flexibility to Segment Different Structures and Points of Interest

Tools that allow freehand annotation such as MeVisLab, ImageJ and MITK can be adapted to structures with different shapes and make modifications on 2D views of

the generated volumes. ITK-SNAP and region growing algorithms like the ones in 3D Slicer and GeoS depend on the number of seeds for their adaptation to the particular shape features of the organs. All of the selected frameworks are not limited to a particular application and can perform segmentations on different organs and points of interest.

4.2.3.3 User-Friendly Segmentation Method Usage

The semi-automatic segmentation method sought by VISCERAL needs to be easy to apply and must be performed in real time in order to allow for optimal user interaction. Segmentation methods like those used in ImageJ and ITK-SNAP require an initial trial and error user interaction to define the best values of the parameters involved. It can take the users some time to understand the functionality of these parameters when they are not familiar with them, as is likely the case for the annotators of the Gold Corpus. The GeoS tool has a fast, straightforward algorithm that can easily be used by the users, and the default parameters given by the tool can be used without need for modification for most of the structures. Adding seed points in the MeVisLab and 3D Slicer region growing algorithms is also a simple task once it is combined with freehand manual corrections.

4.2.3.4 Optimal Visualization of Segmentation and Medical Images

The ImageJ framework has an independent window visualization that requires the handling of multiple open windows and manual interaction for the user to navigate in 3D medical images. MITK and 3D Slicer have a better visualization of the data with the three views visible at the same time and a multiplanar 3D representation or volume rendering that the user can zoom in and out, rotate and navigate with the mouse. One drawback in MITK is that changing between images can cause losing the defined orientation of the image requiring the user to reset the desired image location for visualization. The GeoS tool has a simple, easy-to-use interface with the three views in which it is possible to make annotations. Unfortunately, volume rendering is still not supported and the segmentations can only be visualized in 2D in each view.

4.2.3.5 Effectiveness of the Segmentation

The purpose of selecting frameworks with semi-automatic segmentation methods is to reduce the amount of work when making the annotations and allowing the radiologists to add their experience and input in the segmentations. Since all of the selected tools are not application oriented to a single type of anatomical structure, they can obtain accurate segmentations with enough user feedback. ImageJ and MeVisLab have the least evolved segmentation methods while ITK-SNAP, 3D Slicer and GeoS

are the best annotation tools to perform semi-automatic segmentations in medical images.

4.2.3.6 Easy Interaction with the User for Corrections in the Semi-automatic Output

Most of the application frameworks contain the option to cut or add new voxels to the segmentation output in 2D slices if the segmentation has leakage or a part is missing. Other tools such as GeoS can improve the segmentations by adding more strokes either in the background or in the foreground from the structure of interest. Once the algorithm is run again, it provides a new segmentation incorporating user input influencing the full 3D volume of the segmentation. This is useful for rapid visual inspection of the results and minor user interaction in any of the views for corrections in the output. Updating small changes however still requires the algorithm to be run fully, even though it has a fast implementation for the whole structure and no freehand correction tool is available in the current GeoS version.

4.2.3.7 Time Required for Complete Structure Segmentation

The GeoS annotation tool is the fastest tool for segmenting a complete structure because of its "lazy annotation" implementation, the good data visualization and the annotation in the three views at the same time. ITK-SNAP and multiple seed algorithms like those in MeVisLab and 3D Slicer can also provide quick segmentations that may need user interaction to correct some leakage or missing parts in the segmented volume. For the other available options, there are efficient algorithms to start the segmentation process but eventually they require manual freehand improvements to refine the segmentations and this can take some time, especially for structures with low intensity contrast and soft edges.

4.2.3.8 Segmentation Method Flexibility to Data Obtained from Different Scanners and Image Contrasts

Some of the segmentation methods that involve thresholding can be very sensitive to image noise and full image contrast of the different structures. A more local definition of the contrast is desired, particularly in MRI where field inhomogeneities are common and intensity values can change significantly even within the same structure. An advantage of the GeoS tool and the ITK-SNAP methods is their local approach that limits the expansion of the segmentations to a specific region. One limitation in the ITK-SNAP method is that it is limited to changes in the image gradient and it can also be affected by image noise and poor soft tissue contrast.

4.2.3.9 Output Format of the Annotations

MeVisLab generates contour segmentation object lists from the segmentation results that have to be converted to a different format for their inclusion in the annotation analysis backend. With 3D Slicer, the user can decide the label of each structure and add subclasses of it. Only in the GeoS tool and the MITK framework, the output of the segmentations can be saved in a NIfTI format without additional plug-ins in accordance with the data format definition for the Gold Corpus annotation for VISCERAL. In addition, the RadLex terminology is expected to already be included in the final version of the GeoS prototype, which makes the annotations comparable and provides a better set-up for long-term use of the annotations.

4.2.3.10 Upgrading of the Tool with Minimum User Support for Its Availability During and After the Project Has Ended

Due to the close collaboration with Microsoft and their interest in supporting medical imaging projects, the GeoS tool was adapted and improved based on various requests. Other frameworks such as MITK, 3D Slicer, and ImageJ can be upgraded using freely available plug-ins. However, their maintenance and specific adaptation for the VISCERAL requirements would have involved significantly more effort from the project consortium.

4.2.4 Selected Software and Technical Aspects

The final decision was made in collaboration with physicians evaluating the considered tools and comparing their usefulness. The Microsoft GeoS annotation tool was selected as the principal annotation tool mainly because of its efficiency and accuracy in the segmentations, which require only a few brush strokes from the user to run segmentations in 3D volumes respecting strong edges. Other advantages over the remaining tools are the tool simplicity and easy-to-use annotation interface with the learning of only a few key presses needed to start using the tool for annotations. 3D Slicer was selected as the secondary annotation tool.

For the organ annotations, in general the GeoS software was used, providing means to an interactive, semi-automated segmentation. Nevertheless, for whole-body MRI (T1w and T2w), where structures such as the vertebrae, the kidneys, the pancreas or certain muscles are only visible from a few coronal slices, the 3D Slicer software was used, since GeoS does not support the annotation of point-like structures. In the following sections, we briefly describe the annotation process in GeoS and 3D Slicer.

Fig. 4.1 Segmentation method: with only a foreground stroke for the kidney (*yellow*) and two background strokes for the surrounding organs (*red*, *left image*), an initial segmentation of the kidney is obtained (*right image*). This segmentation can be further refined with more strokes

4.2.4.1 GeoS

By clicking "LoadVolume" on the initial screen, the user selects the assigned image file to be annotated. The segmentation can be started with only a few strokes in the desired structure (foreground) pressing Shift and a left mouse drag (Fig. 4.1, marked yellow). To limit the extension of the segmentation, strokes outside the structure (background) were created with Shift and right mouse drag (Fig. 4.1, marked red). For better automated segmentation, foreground strokes were put in all three views. The segmentation process is then started. The created segment could then be improved by adding more strokes in the structure and in the background and by running the segmentation process again. Five-to-ten iterations were needed to have a good match of the created segmentation with the anatomical structure in the volume. The segmentation tool has five modifiable parameters: iterations, margin, gamma, pre-smoothing and post-smoothing. For most of the segmentations, a margin of 10 and a gamma of 1 were used. For large organs, such as the liver or the lung, the margin was reduced. For the other parameters, the default presets were used. Finally, the created annotations were saved in separate NIfTI files without modifying the original images. In this way, the annotation for one whole-body volume was completed in 3–4 h. A few examples of segmented annotations from various organs are shown in Fig. 4.2.

4.2.4.2 3D Slicer

The DICOM volumes are loaded into the software and the Annotation Module is used to annotate structures of interest. The landmarks within the patient coordinate system are then saved and exported in text files.

Fig. 4.2 Annotation example of different structures in an abdominal CT

4.3 VISCERAL Ticketing Tool/Framework

Manually annotating organs or landmarks in images is a complex process involving many participants, including the administrators of the process, the experts that do the annotation and the experts responsible for quality control. In the VISCERAL project, a system was developed to simplify the management of the manual annotation process. It is based on the commonly used process from software engineering of assigning *tickets* to people for tasks that should be done, where the status of the completion of each assigned task can be tracked. For medical image annotation, an *annotation ticket* is assigned to an annotator requesting the segmentation of a specific organ or identification of landmarks in a specific image. The VISCERAL ticketing framework is designed to monitor and manage the full life cycle of an annotation ticket, to provide an interface for annotators and quality control (QC) team members for ticket submission. The framework consists of three main components:

1. Ticketing Database: A MySQL database that stores information of volumes to annotate, annotators, annotation types, tickets and their states (pending, submitted, QC passed and QC failed).
2. Backend: A backend implemented in MATLAB, to manage volumes, annotators, ticket types and annotation tickets. The backend is used to distribute tickets, to perform automated quality checks and to distribute QC tickets of submitted annotations.
3. Frontend: A Web interface that is used by annotators and QC team members to receive their assigned tickets and to submit annotations and QC results.

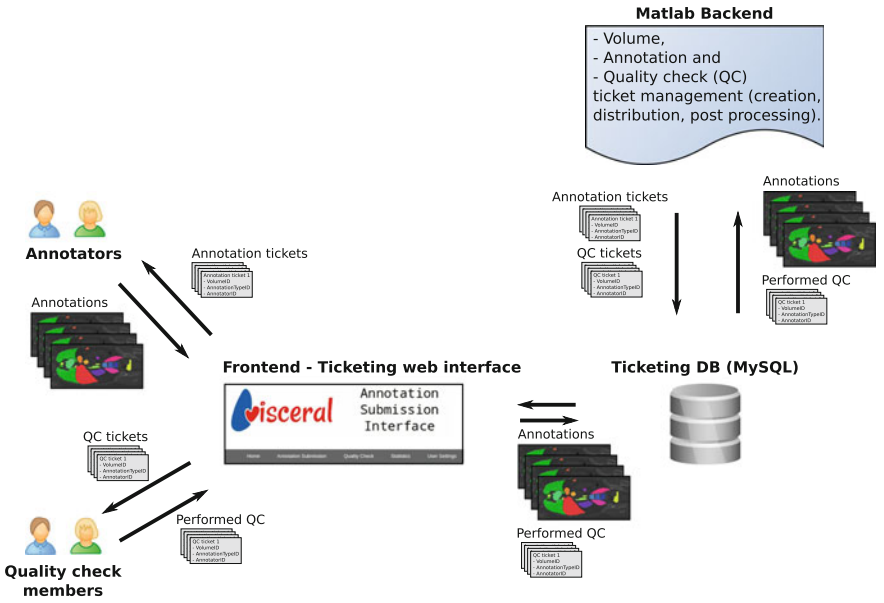

Fig. 4.3 Workflow overview of the VISCERAL ticketing framework

Figure 4.3 provides an overview of the ticketing system implemented within the VISCERAL project. The framework source code is available,[10] which provides create statements for the ticketing DB, the source of the Web interface, getter and setter functions of the backend implemented in Matlab including a tutorial and installation guidelines to set up the framework. The interface is designed using Java.

4.3.1 Ticketing System Database

The database (DB) of the ticketing system is created by SQL scripts provided in the ticketing repository. All relevant information is stored in five DB tables:

- **Annotator**: Identified by an AnnotatorID, holds next to contact information, name and password (for login) a flag indicating if the annotator is currently available and an additional flag if the annotator is considered a QC team member.
- **Volume**: A volume is identified by its PatientID and VolumeID. Additionally, the modality, body region and the filename are stored.
- **AnnotationType**: Entries in this table define which types of annotations can be managed by the ticketing system. Each entry is identified by its AnnotationTypeID. Additionally, the name, the file extension of the submitted files, the remote upload

[10] https://github.com/Visceral-Project/annotationTicketingFramework.

directory, the category (segmentation, landmarks,...) and an optional string describing the files prefix are stored. Exemplary entries of this table are created within the given SQL scripts.
- **Status**: Defines all states a ticket can have during its life cycle. A status is identified by its StatusID and stores its name and description as well as to which type of annotators (QC and normal annotators) the status option is available in the ticketing Web interface. Default entries of this table are created within the given SQL scripts.
- **Annotation**: This table represents an annotation ticket. An annotation entry is identified by its PatientID and VolumeID, the AnnotationTypeID, the StatusID and the AnnotatorID of the annotator to whom a ticket is assigned. Additionally, the filename, a timestamp, the ID of the annotator who performs the QC of the ticket and a QC comment are stored for each ticket.

Figure 4.4 illustrates the ER (entity relationship) diagram of the resulting database.

4.3.2 Annotation Ticket Life Cycle

The typical life cycle of an annotation ticket within the VISCERAL project can be outlined as follows:

1. Creation of an annotation list with the annotations that need to be done (tickets) and its upload to the Web interface.
2. The Web interface has a login user name and password for each of the annotators.
3. The annotator ID is used in the naming of the tickets:
   subjectXX acquisitionZZ[modalityYY] RadLexID annotatorID:nii
4. The annotators upload their files next to the ticket and the name of the file is implemented to be the same as the ticket for their backend analysis.
5. All the annotations are saved in the same folder to download them and use them in the analysis.
6. The annotation backend produces a new list of tickets for the new annotations needed and from which annotators.
7. The list in the interface is updated.
8. Depending on the type of annotation, an automated quality check is performed to detect common annotation errors, such as empty label volumes, incorrect file extensions or wrong naming of landmarks.
9. If the annotation passes the automated quality check, the ticket is assigned to a quality control (QC) annotator; otherwise, it is reassigned to the annotator for corrections to the annotation.
10. The QC annotator receives the QC ticket through the Web interface, performs the QC and submits the QC result (including textual feedback if the QC is negative) through the Web interface.

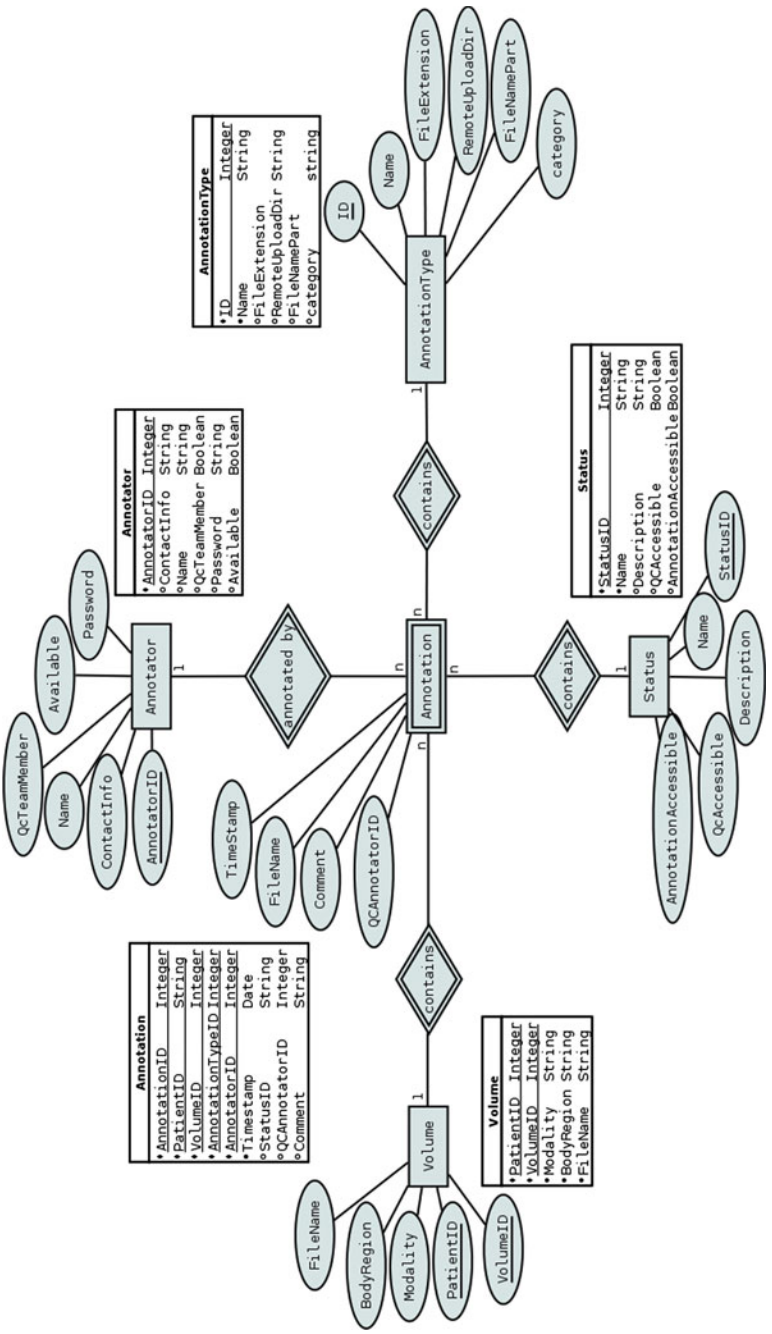

Fig. 4.4 Entity relationship diagram of the VISCERAL Ticketing System DB

11. The ticket reaches its final state if the QC is positive; otherwise, the annotator receives textual feedback and the ticket is reassigned for annotation.
12. The annotators receive their new set of tickets when they login.

4.3.3 Manual Annotation Instructions

In order to ensure reproducible annotations from multiple annotators, detailed instructions on the annotation of each organ were created. These annotation instructions describe the specificities of performing the anatomical structure segmentations and landmark locations of the Gold Corpus, complete with illustrations. To reduce the probability of misunderstandings, the instructions were written in the native language of the radiologists doing the annotation (in this case Hungarian). Below, we describe some aspects that need to be made explicit in the annotation instructions and show some of the example images.

4.3.3.1 Organ Segmentation

When delineating the organs, we face the problem of defining the outer extensions of the structure, requiring a definition of what part of a connected structure is still

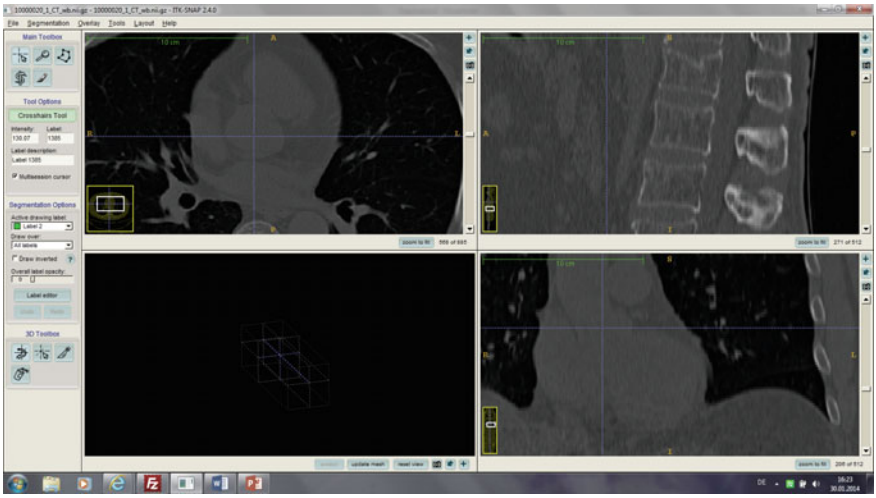

Fig. 4.5 Instructions for segmenting the aorta beginning in whole-body CTs: in the above *left* window, an axial slice is shown and the cross wires are centred at the aortic bulb in the height of the aortic valve. The below *right* window shows a coronary slice. The cross wire is located in the aortic bulb. You should segment until you see the diameter of aorta in the axial slices being in the region of aortic bulb and you could control that on the coronary slice

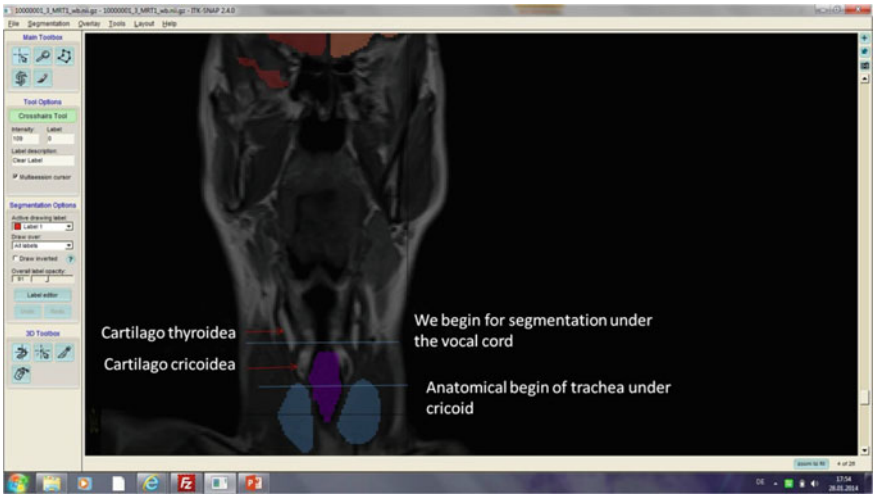

Fig. 4.6 Coronal whole-body MRI T1 weighted of the head/neck demonstrating the beginning of the trachea: the anatomical beginning of the trachea is under the cricoids, but that is not suitable for manual segmentation, so we define the beginning of trachea for segmentation under the vocal cord. Trachea (*purple*) and thyroid gland (*blue*)

"within" the organ and what is already "outside" of the organ, belonging to a different structure (that may or may not be in the list of annotated organs). Some organs, such as the lungs, liver and kidneys, have a hilum, which is a depression or indentation where vessels and nerves enter. It must be explicitly specified how to handle the hilum in the manual annotation — we specified that the hilum has to be cut off during the segmentation process. It is also often useful to provide an "algorithm" for the annotator to follow, as illustrated in Figs. 4.5 and 4.6. Some examples of organ segmentations are shown in Fig. 4.7.

4.3.3.2 Landmark Location

Anatomical landmarks are the locations of selected anatomical structures that can be identified in images of multiple modalities, such as CT or MRI, unenhanced or enhanced scans, whole-body images or with limited field of view. Their universal nature makes them important as a first step in parsing image content, or for triangulating other more specific anatomical structures. Overall, 40 landmarks were selected to be identified. Examples of annotated landmarks are shown in Fig. 4.8.

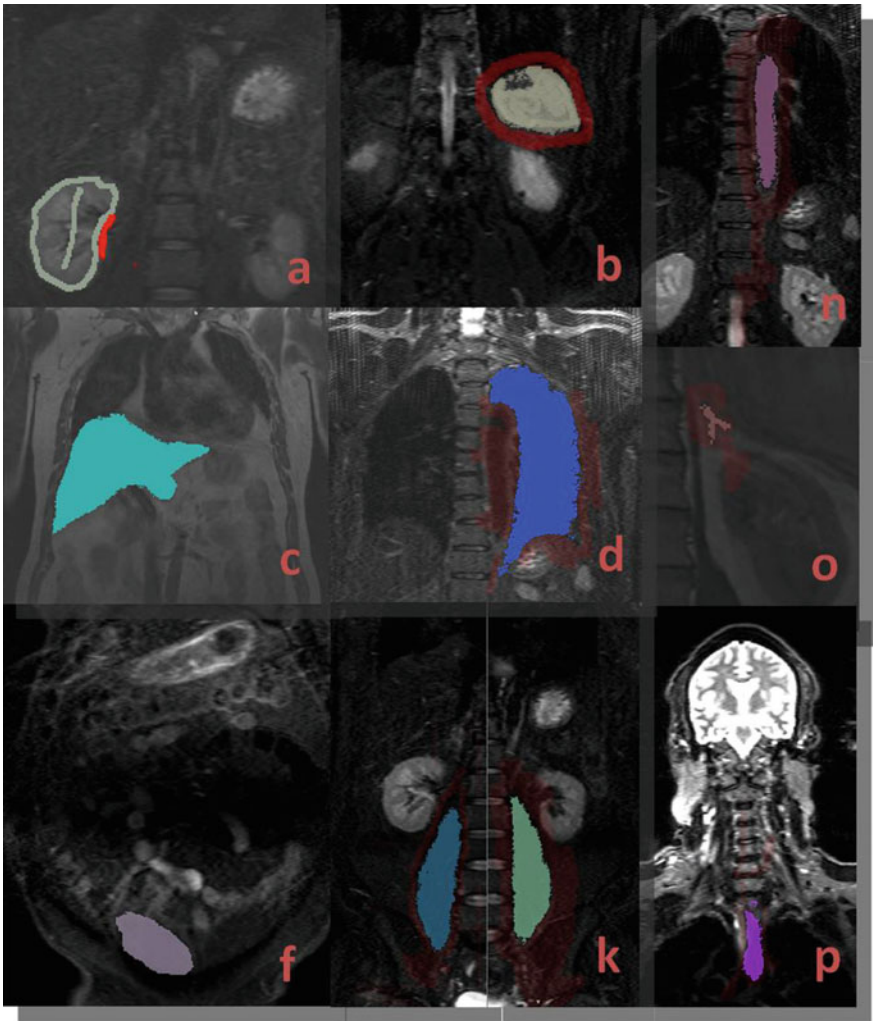

Fig. 4.7 Examples of organ segmentation in MRI T1w and T1w. **a** Kidney: Marked *grey* is the foreground stroke for the liver. Marked *red* is the background stroke defining the border of the organ. **b** Spleen. **c** Liver. **d** Lung (*left*). **f** M. rectus abdominis (*right*). **k** M. psoas major. **n** Aorta. **o** Adrenal gland. **p** Trachea

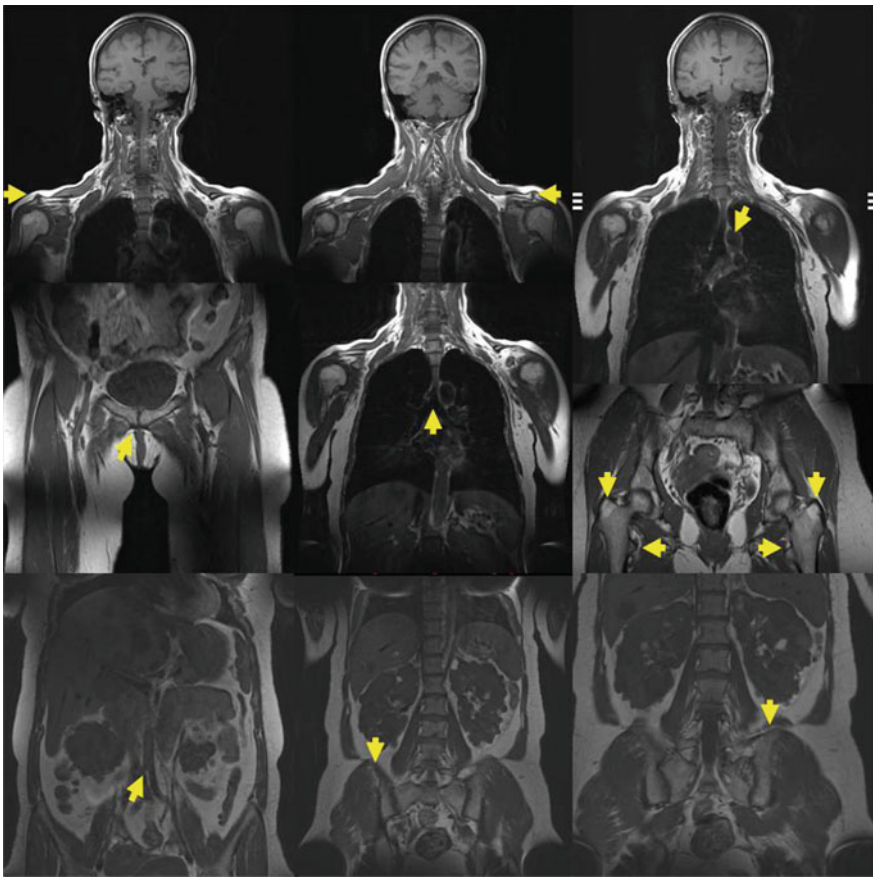

Fig. 4.8 Landmark examples. From *right top* to *left bottom*: *right* and *left* lateral end of the clavicle, tip of aortic arch, symphysis below, tracheal bifurcation, trochanter major at the tip and trochanter minor (most medal part), aortic bifurcation, Crista iliaca (at the *top*)

4.4 Inter-annotator Agreement

To analyse the reliability of Gold Corpus annotations, the agreement of different experts was investigated by comparing multiple annotations of a specific structure obtained from multiple annotators. The similarity of two annotations for this purpose is measured using the Dice coefficient [5], which is a spatial overlap measure that is 1 for perfect overlap and 0 if no overlap of two segmentations is present. The Dice coefficient, also called the overlap index, is the most frequently used metric in validating medical volume segmentations. Zou et al. [15] used the Dice coefficient as a measure of the reproducibility as a statistical validation of manual annotation where segmentors repeatedly annotated the same image [13]. In this context, inter-annotator agreements are reported independently for each structure in each modality

Table 4.2 Inter-annotator agreements of structures and modalities addressed within VISCERAL

Structure	CT Wb	CT-ce Thx/Abd	MRT1 Wb	MRT1 cefs-Abd
Lung R	0.974	0.946	0.925	
Lung L	0.971	0.945	0.902	
Kidney R	0.889	0.937	0.917	0.908
Kidney L	0.921	0.929	0.909	0.865
Liver	0.950	0.965	0.891	0.932
Spleen	0.946	0.934	0.685	0.925
Urinary Bladder	0.875	0.933	0.842	0.819
Psoas Major R	0.840	0.854	0.836	0.823
Psoas Major L	0.847	0.848	0.849	0.802
Trachea	0.894	0.877	0.768	
Aorta	0.884	0.856	0.849	0.768
Sternum	0.891	0.810		
1st Lumbar Vertebra	0.811	0.914	0.744	0.546
Muscle Body of Rectus Abdominis R	0.811	0.709		
Muscle Body of Rectus Abdominis L	0.734	0.637		
Pancreas	0.615	0.785	**0.486**	0.639
Gallbladder	0.689	0.857	0.742	
Thyroid Gland	0.658	0.781		
Adrenal Gland R	**0.347**	0.671	**0.465**	**0.305**
Adrenal Gland L	**0.485**	0.743	0.545	**0.338**

of the Gold Corpus and are obtained by comparing Gold Corpus segmentations from different annotators to additionally performed segmentations of the same structures in the same volumes (double annotations). Inter-annotator agreement is finally derived by averaging the Dice coefficients of all double annotations performed for a specific structure and modality and are shown in Table 4.2. Missing values are due to the structure being out of field such as the trachea in MRT1cefs-abdominal volumes (MRT T1 weighted contrast-enhanced sequence with fat saturation of the abdomen) or bad contrast of the addressed structure in a certain modality.

Organs such as the adrenal glands show, depending on the annotated modality, a Dice coefficient smaller than 0.5 (see the bold values in Table 4.2). The reason for this is probably that the adrenal glands have the best contrast in contrast-enhanced CT compared to other sequences. The CT volumes have overall the best average Dice coefficients also in small anatomical structures such as the adrenal glands. The adrenal glands, the thyroid gland, the pancreas and the bodies of the rectus abdominis muscles have the smallest average Dice coefficients which are smaller and/or equal

than 0.8 in all four modalities: in whole-body CT and MRI (T1 weighted), in contrast-enhanced CT of thorax and abdomen and in contrast-enhanced and fat-saturated T1-weighted MRI. The contrast-enhanced CT sequence is the best for the adrenal glands, the thyroid gland and the pancreas. The reason for that is a worse contrast in the other sequences at the adrenal gland regions, especially at the right one in a small window between liver, right kidney and vertebral column in the fatty tissue. For the bodies of the rectus abdominis muscle, the native whole-body CT sequence is ahead, probably due to a better contrast for this structure without contrast media. If the muscle bodies of the rectus abdominis are small, the contrast between fatty tissue and muscle is also not sufficient to reliably and repeatably segment these structures. The difference of the pancreatic tissue and the fat-surrounding tissue is not high enough without contrast media and therefore difficult to annotate. Figure 4.9 visualizes the agreement between two annotators based on a liver segmentation in a contrast-enhanced CT volume of the thorax and abdomen and shows that for this structure the discrepancy is only marginal.

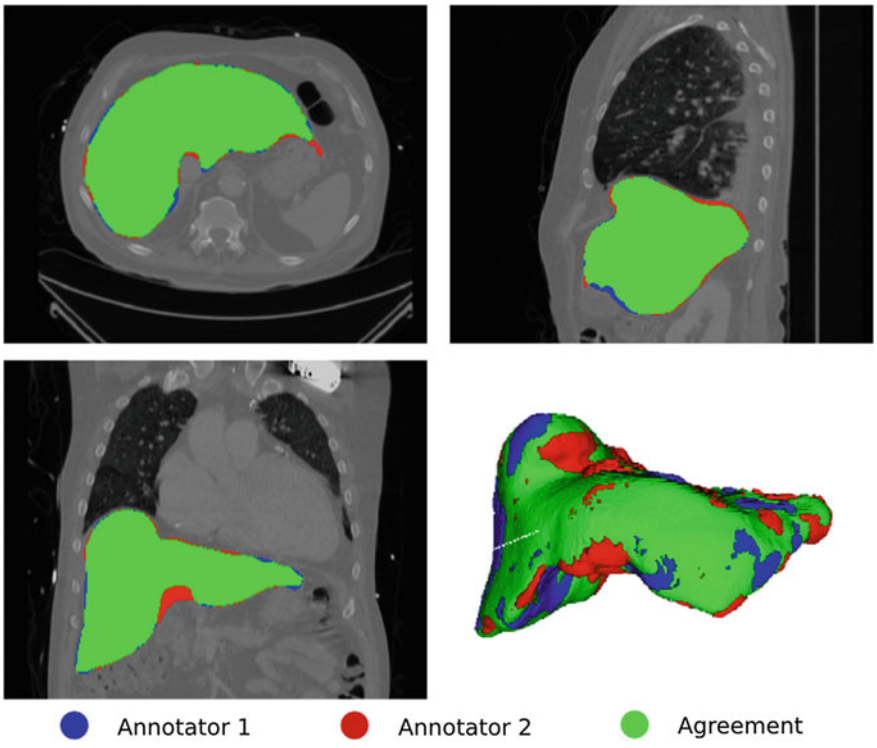

Fig. 4.9 Visualization of the agreement of two annotators based on a liver segmentation in a contrast-enhanced CT volume of the thorax and abdominal region

4.5 Conclusion

This chapter described the annotation of medical images that was performed in the VISCERAL project. Both the selection of the tool to annotate the 3D images and the quality management of the annotations are extremely important. A good choice of an annotation tool can significantly limit the amount of work required to do the annotations. Semi-automatic solutions like the ones chosen allow to rapidly achieve good segmentation results.

The quality management also showed that this process can be automated to optimize the outcomes. Not only is a detailed description of the structures to annotate important, but also regular controls and manual checks to compare the written description with the actual practice. There is always subjectivity in annotations, so double annotations are essential to judge the subjectivity of a task. However, only if the control is systematic can the subjectivity be limited and estimated well.

Acknowledgements The research leading to these results has received funding from the European Union Seventh Framework Programme (FP7/2007–2013) under grant agreement 318068 (VISCERAL).

References

1. Bitter I, Van Uitert R, Wolf I, Ibanez L, Kuhnigk JM (2007) Comparison of four freely available frameworks for image processing and visualization that use ITK. IEEE Trans Vis Comput Graph 13(3):483–493
2. Caban JJ, Joshi A, Nagy P (2007) Rapid development of medical imaging tools with open-source libraries. J Digit Imaging 20(1):83–93
3. Chapman BE, Wong M, Farcas C, Reynolds P (2012) Annio: a web-based tool for annotating medical images with ontologies. In: 2012 IEEE second international conference on healthcare informatics, imaging and systems biology. IEEE, New Jersey, p 147
4. Criminisi A, Sharp T, Blake A (2008) GeoS: geodesic image segmentation. In: Forsyth D, Torr P, Zisserman A (eds) ECCV 2008. LNCS, vol 5302. Springer, Heidelberg, pp 99–112. doi:10.1007/978-3-540-88682-2_9
5. Dice LR (1945) Measures of the amount of ecologic association between species. Ecology 26(3):297–302
6. Doi K (2005) Current status and future potential of computer-aided diagnosis in medical imaging. Br J Radiol 78:3–19
7. Engle RL (1992) Attempts to use computers as diagnostic aids in medical decision making: a thirty-year experience. Perspect Biol Med 35(2):207–219
8. Hanbury A, Müller H, Langs G, Weber MA, Menze BH, Fernandez TS (2012) Bringing the algorithms to the data: cloud-based benchmarking for medical image analysis. In: Catarci T, Forner P, Hiemstra D, Peñas A, Santucci G (eds) CLEF 2012. LNCS, vol 7488. Springer, Heidelberg, pp 24–29. doi:10.1007/978-3-642-33247-0_3
9. Langs G, Hanbury A, Menze B, Müller H (2013) VISCERAL: towards large data in medical imaging — challenges and directions. In: Greenspan H, Müller H, Syeda-Mahmood T (eds) MCBR-CDS 2012. LNCS, vol 7723. Springer, Heidelberg, pp 92–98. doi:10.1007/978-3-642-36678-9_9

10. Mata C, Oliver A, Torrent A, Marti J (2012) Mammoapplet: an interactive Java applet tool for manual annotation in medical imaging. In: 2012 IEEE 12th international conference on bioinformatics and bioengineering (BIBE). IEEE, New Jersey, pp 34–39
11. Pieper S, Halle M, Kikinis R (2004) 3D slicer. In: IEEE international symposium on biomedical imaging: nano to macro. IEEE, New Jersey, pp 632–635
12. Sharma N, Aggarwal LM (2010) Automated medical image segmentation techniques. J Med Phys Assoc Med Phys India 35(1):3
13. Jiménez-del Toro OA, Müller H, Krenn M, Gruenberg K, Taha AA, Winterstein M, Eggel I, Foncubierta-Rodríguez A, Goksel O, Jakab A, Kontokotsios G, Langs G, Menze B, Salas Fernandez T, Schaer R, Walleyo A, Weber MA, Dicente Cid Y, Gass T, Heinrich M, Jia F, Kahl F, Kechichian R, Mai D, Spanier AB, Vincent G, Wang C, Wyeth D, Hanbury A (2016) Cloud-based evaluation of anatomical structure segmentation and landmark detection algorithms: visceral anatomy benchmarks. IEEE Trans Med Imaging 35(11):2459–2475
14. Yushkevich PA, Piven J, Hazlett HC, Smith RG, Ho S, Gee JC, Gerig G (2006) User-guided 3D active contour segmentation of anatomical structures: significantly improved efficiency and reliability. Neuroimage 31(3):1116–1128
15. Zou KH, Warfield SK, Bharatha A, Tempany CM, Kaus MR, Haker SJ, Wells WM III, Jolesz FA, Kikinis R (2004) Statistical validation of image segmentation quality based on a spatial overlap index[1]: scientific reports. Acad Radiol 11(2):178–189